General Multilevel Models (MLMs) for Two-Level Nested Data: Random Slopes and Cross-Level Interactions

- Topics:
  - Random effects of level-1 predictors
  - Cross-level interactions
  - Fixed, random, and systematically varying effects
  - Model-predicted variance-covariance matrices
  - Fun with maximum likelihood estimation
  - Missing data in MLMs
  - Predicting level-2 unit-specific random effects

Fixed Effects of Predictors in MLM

- **Fixed effects of level-2 between-group predictors:**
  - Level-2 (BG) main effects reduce level-2 (BG) random intercept variance
  - Level-2 (BG) interactions also reduce level-2 (BG) random intercept variance

- **Fixed effects of level 1 predictors:**
  - Level-1 (WG) main effects reduce Level-1 (WG) residual variance
  - Level-1 (WG) interactions also reduce Level-1 (WG) residual variance

- **Effect on level-2 random intercept variance by type of level-1 predictor:**
  - If the level-1 predictor has level-2 variance (e.g., Grand-MC predictors), then the implied level-2 effect will also affect the level-2 random intercept variance
    - Smushed effects in same direction will reduce level-2 intercept variance, but smushed effects in the opposite direction can actually increase level-2 intercept variance instead
  - If the level-1 predictor DOES NOT have level-2 variance (e.g., Group-MC predictors), then its reduction in the level-1 residual variance will cause an INCREASE in level-2 random intercept variance
    - Same thing happens with Grand-MC level-1 predictors, but you don't generally see it
    - It's just an artifact that the estimate of true random intercept variance is:

True $\tau^2_{U_0}$ = observed $\tau^2_{U_0} - \frac{\sigma^2_e}{n}$ → so if only σ^2_e decreases, $\tau^2_{U_0}$ increases

Fixed and Random Effects of Level-1(Person-Level) Predictors

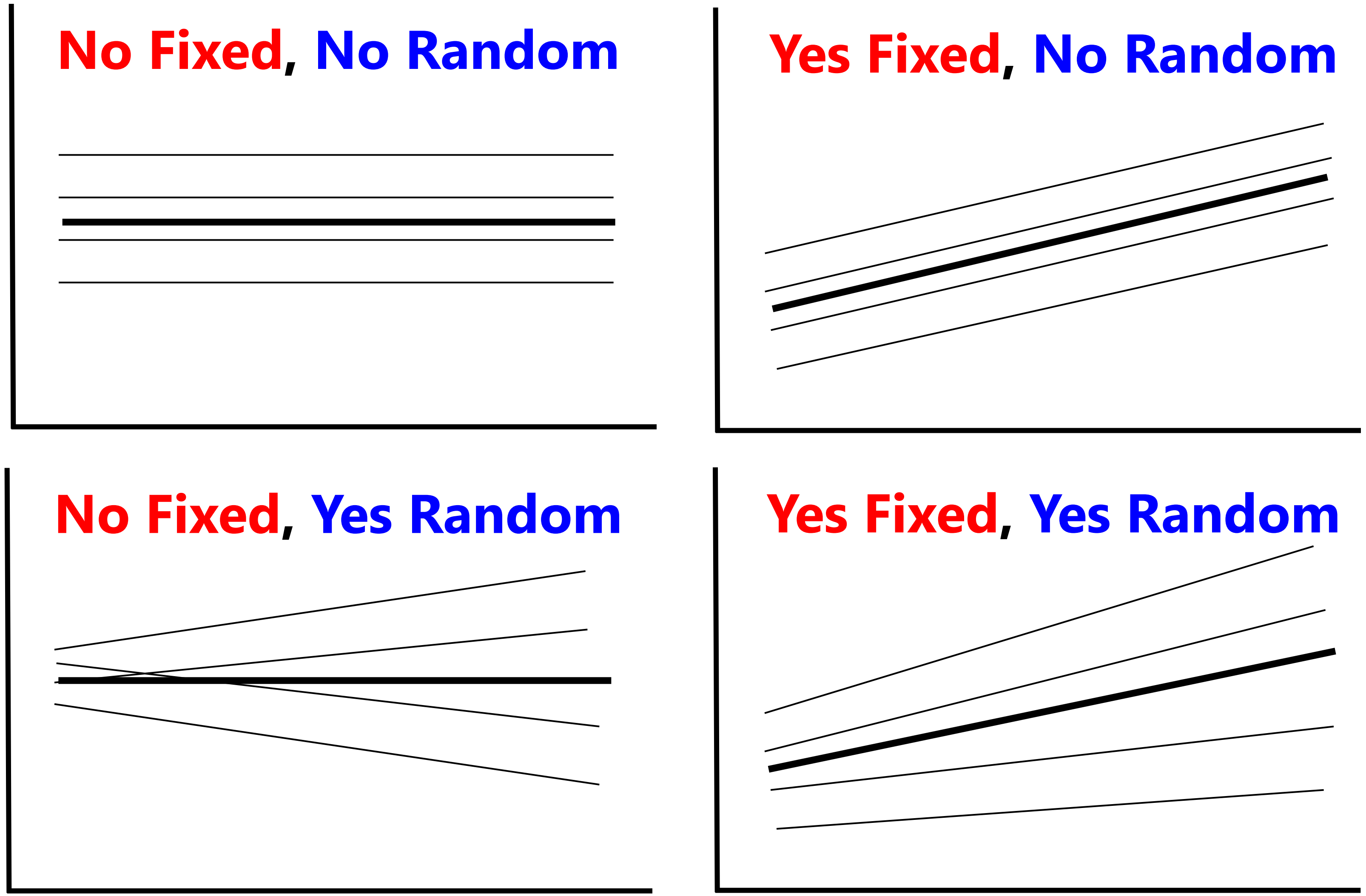
- 2 questions about possible effects of *level-1 (person-level) predictors*:

1. **Is there a predictor effect on average?**
   - Is its regression line not flat?
   - Significant **FIXED** effect—this is what we've done so far

2. **Does the predictor effect vary across groups?**
   - Does each group need their own regression line?
   - Significant **RANDOM** effect—this is a new concept

Fixed and Random Effects of Level-1 x_{pg}

(Note: The group intercept is random in every figure)



Adding Random Slopes to a Grand-MC Level-1 Model

x_{pg} is grand-mean-centered into $L1x_{pg}$, WITH GMx_g at L2:

Level 1: $y_{pg} = \beta_{0g} + \beta_{1g}(L1x_{pg}) + e_{pg}$

Level 2: $\beta_{0g} = \gamma_{00} + \gamma_{01}(GMx_g) + U_{0g}$

$\beta_{1g} = \gamma_{10} + U_{1g}$

$L1x_{pg} = x_{pg} - C_1$ → it still has both Level-2 BG and Level-1 WG variation

$GMx_g = \bar{X}_g - C_2$ → it has only Level-2 BG variation

U_{1g} is a random slope for the WG effect of x_{pg}

γ_{10} becomes the WG effect → *unique* level-1 effect after controlling for GMx_g

γ_{01} *becomes* the contextual effect that indicates <u>how the BG effect differs from the WG effect</u> → *unique* level-2 effect after controlling for $L1x_{pg}$ → does group matter beyond individuals?

Adding Random Slopes to a Grand-MC Level-1 Model: Example

- For example: level-1 students (p) nested in level-2 schools (g), level-1 x_{pg} = time parents read to them, y_{pg} = end-of-grade reading test
- **Level-1:** $\textbf{Test}_{pg} = \beta_{0g} + \beta_{1g}(\textbf{Read}_{pg} - C_1) + e_{pg}$
- **Level-2:** $\beta_{0g} = \gamma_{00} + \gamma_{01}(\overline{\textbf{Read}}_{g} - C_2) + U_{0g}$

  $\beta_{1g} = \gamma_{10} + U_{1g}$

- γ_{00} = fixed intercept: expected test score for a student with reading = C_1 from a school with school mean reading = C_2
- γ_{10} = fixed L1 Within-Group slope: difference in test score per one-unit higher in student (within-school) reading than others in your school
- γ_{01} = fixed L2 contextual slope: incremental difference in test score per one-unit higher school mean reading than other schools (controlling for student read)
- $\gamma_{01} + \gamma_{10}$ = fixed L2 Between-Group slope: difference in test score per one-unit higher in school mean reading than other schools (NOT controlling for student reading)

Adding Random Slopes to a Grand-MC Level-1 Model: Example

- **Level-1:** $\textbf{Test}_{pg} = \beta_{0g} + \beta_{1g}(\mathbf{Read}_{pg} - C_1) + e_{pg}$
- **Level-2:** $\beta_{0g} = \gamma_{00} + \gamma_{01}(\overline{\mathbf{Read}}_{g} - C_2) + U_{0g}$

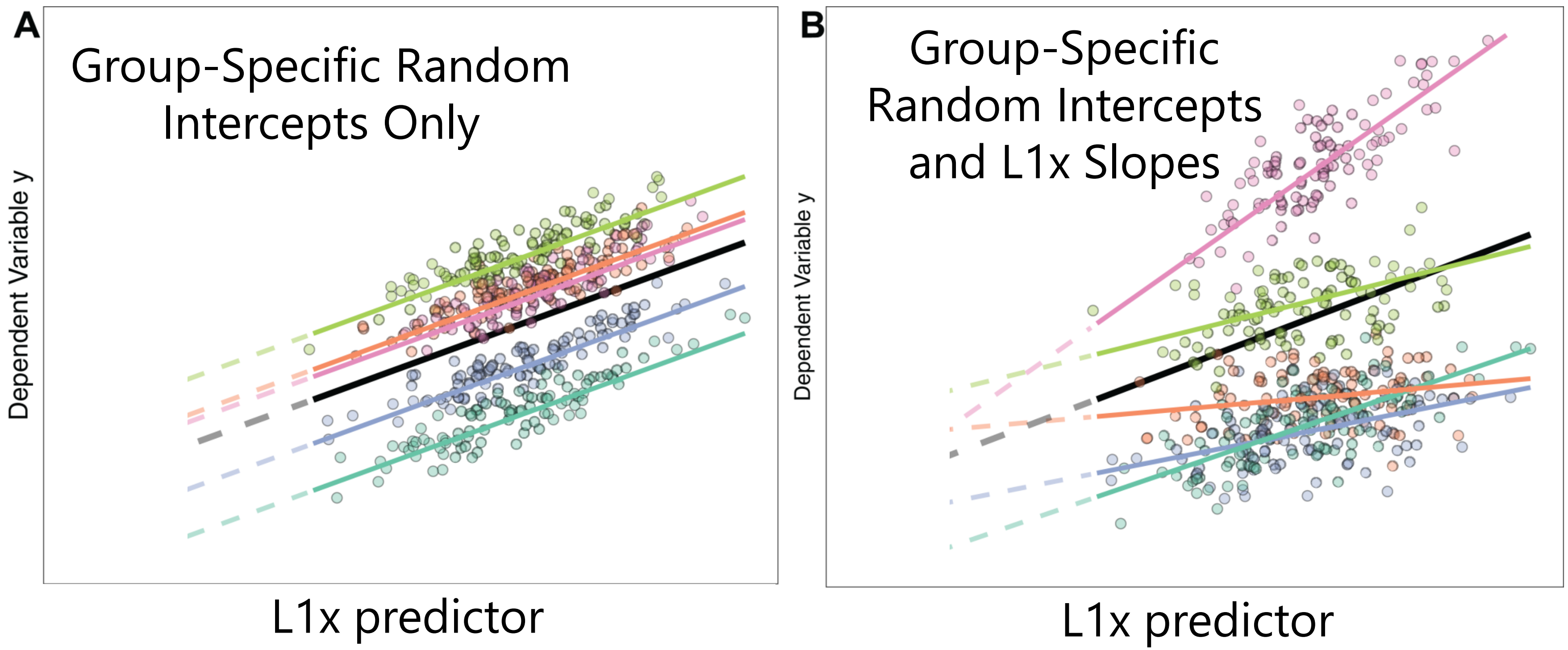
  $\beta_{1g} = \gamma_{10} + U_{1g}$

- $\mathbf{U_{0g}}$ = level-2 random intercept = deviation between actual and school mean test score conditional on (predicted from) $\gamma_{00} + \gamma_{01}(\overline{\mathbf{Read}}_{g} - C_2)$ for school *g* **at** $C_1 = 0$ → for unknown reasons, so is "error"

- $\mathbf{U_{1g}}$ = level-2 random slope for effect of Read_{pg} → deviation between the sample mean slope (γ_{10}) and the slope for school *g*
  → Is <u>unconditional: ALL the unknown reasons</u> why some schools have larger student reading effects on scores than others in same school
  → Although not shown here, we also add a covariance with $\mathbf{U_{0g}}$
  → This means that $\mathbf{U_{0g}}$ is now specific to when Read_{pg} = 0 (C_1 here)

- $\mathbf{e_{pg}}$ = level-1 residual = deviation of predicted from student *p*'s real test score

Random Level-1 Slopes Across Groups



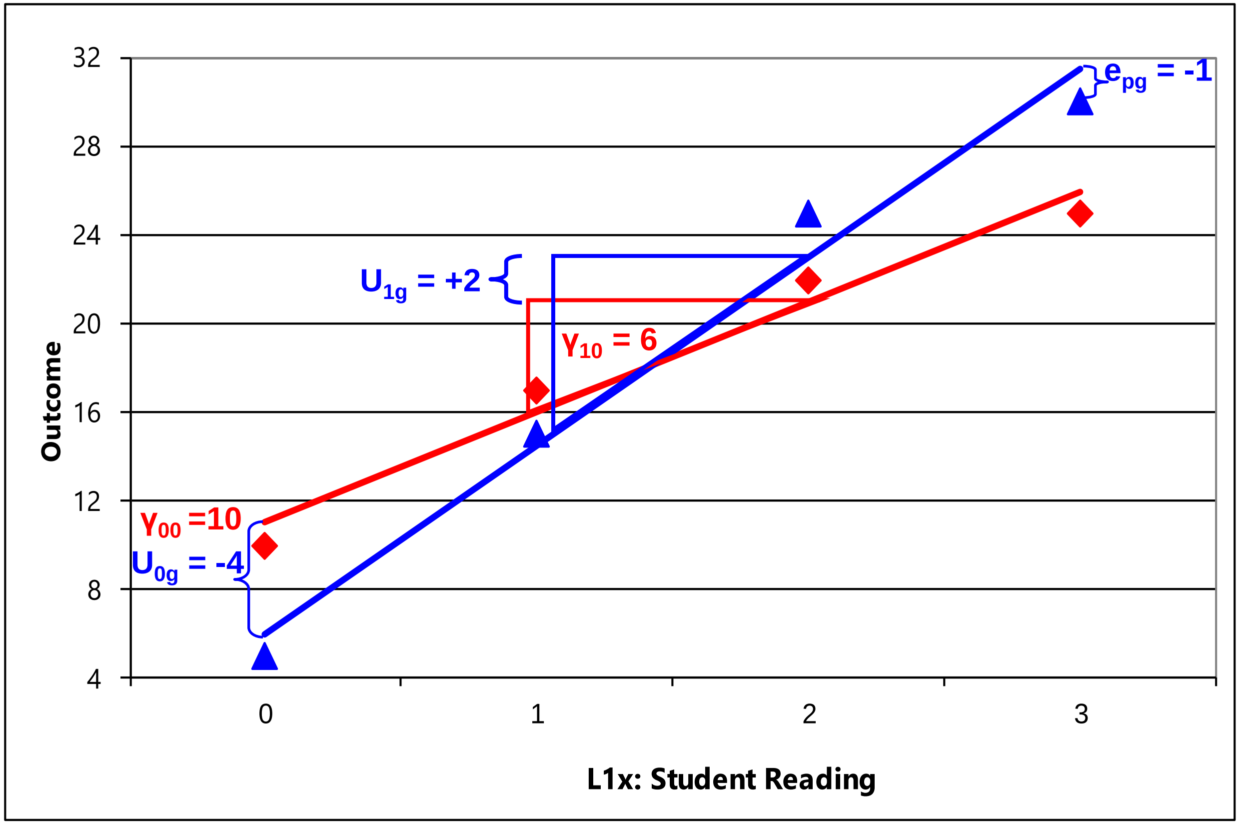
- Both: the black line conveys the fixed slope for L1x, γ_{10}
- Right: deviation for each group's L1x slope is given by U_{1g}
  - Left: $\beta_{1g} = \gamma_{10}$ Right: $\beta_{1g} = \gamma_{10} + U_{1g}$

Image borrowed from: https://peerj.com/articles/4794/

Random Slope Model (Holding constant $\overline{\text{Read}_g} - C_2$)

$$y_{pg} = [\gamma_{00} + \gamma_{01}(\overline{\text{Read}_g} - C_2) + U_{0g}] + [\gamma_{10} + U_{1g}](\text{Read}_{pg} - C_1) + e_{pg}$$

- γ_{00}: Fixed Intercept
- γ_{01}: Fixed L2 Slope
- U_{0g}: Random **Intercept** Deviation
- γ_{10}: Fixed L1 Slope
- U_{1g}: Random **Slope** Deviation
- e_{pg}: error for person p in group g



7 Parameters:

2 Fixed Effects:
γ_{00} Int, γ_{01} & γ_{10} slopes

4 Parameters in the Model for the Variance:

L2 U_{0g} Random Intercept Variance = $\tau^2_{U_0}$

L2 U_{1g} Random Slope Variance = $\tau^2_{U_1}$

L2 Random Int-Slope Covariance = $\tau_{U_{01}}$

L1 e_{pg} Residual Variance = σ^2_e

Relative Model Fit by Side of the Model

- **Nested models** (i.e., in which one is a subset of the other) can differ from each other in **two distinct ways**

- **Model for the Means** → which fixed effects of predictors should be included in the model
  - Significance tests for whether new fixed effects improve the model are done using univariate (DF_{num}=1) or multivariate (DF_{num}>1) Wald tests

- **Model for the Variance** → which level-1 fixed effects also need level-2 random effect variances
  - Significance tests must be done via **relative model fit using −2LL**
  - Cannot use univariate Wald test *p*-values for testing significance of variances because those *p*-values use a two-sided sampling distribution for what the variance could be (but since variances cannot be negative, those *p*-values are not valid)

Statistical Significance of Fixed Effects: What letters will I get?

Fixed effects can be tested via **Wald** tests: the ratio of its estimate/SE forms a statistic we compare to a distribution

	Denominator DF is assumed infinite	Denominator DF is estimated instead
Numerator DF = 1	use z distribution (Mplus, STATA)	use t distribution (SAS, SPSS)
Numerator DF > 1	use χ^2 distribution (Mplus, STATA)	use F distribution (SAS, SPSS)
Denominator DF (DDFM) options	not applicable, so DDF is not given	SAS: BW and KR SAS and SPSS: Satterthwaite Stata 14: BW, Satt, and KR

Denominator DF (DDF) Methods for Tests of Fixed Effects

- **Between-Within** (DDFM=BW in SAS, not in SPSS):
  - Total DDF (T) comes from total number of observations, separated into level-2 for *N* persons and level-1 for *n* occasions
    - **Level-2 DDF** = *N* – #level-2 fixed effects
    - **Level-1 DDF** = Total DDF – Level-2 DDF – #level-1 fixed effects
    - Level-1 effects with random slopes still get level-1 DDF
- **Satterthwaite** (DDFM=Satterthwaite in SAS, default in SPSS):
  - More complicated, but analogous to two-group *t*-test given unequal residual variances and unequal group sizes
  - Incorporates contribution of variance components at each level
    - Level-2 DDF will resemble Level-2 DDF from BW
    - Level-1 DDF will resemble Level-1 DDF from BW if the level-1 effect is not random, but will resemble level-2 DDF if it is random

Denominator DF (DDF) Methods for Tests of **Fixed Effects**

- **Kenward-Roger** (DDFM=KR in SAS, not in SPSS):
  - Adjusts the sampling covariance matrix of the fixed effects and variance components to reflect the uncertainty introduced by using large-sample techniques of ML/REML in small *N* samples
  - This creates different (larger) SEs for the fixed effects
  - Then uses Satterthwaite DDF, new SEs, and *t* to get *p*-values

- In an unstructured variance model, all effects use level-2 DDF
  - Only relevant in balanced longitudinal data
- Differences in inference not likely to matter often in practice
  - e.g., critical *t*-value at DDF=20 is 2.086, at infinite DDF is 1.960
- When in doubt, use KR (is overkill at worst, becomes Satterthwaite)
  - I used Satterthwaite in the book to maintain comparability across programs

Comparing Models for the Variance

- Testing **random effects** requires assessing **relative model fit**: how well does the model fit relative to other possible models?

- Relative fit is indexed by overall model **log-likelihood** (**LL**):
  - Starts with log of likelihood for each group's outcomes given estimates of model parameters (stay tuned for estimation details)
  - Sum log-likelihoods across all independent groups = **model LL**
  - Two flavors: Maximum Likelihood (ML) or Restricted ML (REML)

- What you get for this on your output varies by software…

- Given as −2*log likelihood (−2LL) in SAS or SPSS MIXED: **−2LL** gives BADNESS of fit, so **smaller** value = better model

- Given as just log-likelihood (LL) in STATA MIXED and Mplus: **LL** gives GOODNESS of fit, so **bigger** value = better model

Comparing Models for the Variance

- **Nested models are compared using their deviance values: −2ΔLL Test** (i.e., Likelihood Ratio Test, Deviance Difference Test)

1. Calculate −2ΔLL: $(-2LL_{fewer}) - (-2LL_{more})$
2. Calculate Δdf: $(\#\ Parms_{more}) - (\#\ Parms_{fewer})$
3. Compare −2ΔLL to χ^2 distribution with df = Δdf
   CHIDIST in excel gives exact p-values for the difference test; so will STATA LRTEST

1. & 2. must be positive values!

- **Add** parameters? Model fit can be **BETTER** (signif) or **NOT BETTER**
- **Remove** parameters? Model fit can be **WORSE** (signif) or **NOT WORSE**

- Nested or non-nested models can also be compared by **Information Criteria** that also reflect model parsimony
  - No significance tests or critical values, just "smaller is better"
  - **AIC** = Akaike IC = −2LL + 2 *(#parameters)
  - **BIC** = Bayesian IC = −2LL + log(N)*(#parameters)
  - What "parameters" means depends on flavor (except in STATA):
    - ML = ALL parameters; REML = variance model parameters only

ML vs. REML in a nutshell

Remember "population" vs. "sample" formulas for calculating variance?

"Population"

$$\frac{\Sigma(y_i - y_{pred})^2}{N}$$

"Sample"

$$\frac{\Sigma(y_i - y_{pred})^2}{N - k}$$

All comparisons must have same *N*!!!	**ML**	**REML**
To select, type...	METHOD=ML (-2 log likelihood)	METHOD=REML *default* (-2 res log likelihood)
In estimating variances, it treats **fixed effects** as...	**Known** (df for having to also estimate fixed effects is not factored in)	**Unknown** (df for having to estimate fixed effects is factored in)
So, in small samples, **L2 variances** will be...	**Too small** (less difference after N=30-50 or so)	**Unbiased** (correct)
But because it indexes the fit of the...	**Entire model** (means + variances)	**Variances model only**
You can compare models differing in...	**Fixed and/or random effects** (either/both)	**Random effects only** (same fixed effects)

Rules for Comparing Models

All observations must be the same across models!

Compare Models Differing In:

Type of Comparison:	**Means Model (Fixed Effects) Only**	**Variance Model (Random Effects) Only**	**Both Means and Variances Model (Fixed and Random)**
Nested? YES, can do significance tests via…	Fixed effect *p*-values from ML or REML -- *OR* -- ML −2ΔLL only (NO REML −2ΔLL)	NO *p*-values REML −2ΔLL (ML −2ΔLL is ok if big N)	ML −2ΔLL only (NO REML −2ΔLL)
Non-Nested? NO signif. tests, instead see…	ML AIC, BIC (NO REML AIC, BIC)	REML AIC, BIC (ML ok if big N)	ML AIC, BIC only (NO REML AIC, BIC)

Nested = one model is a direct subset of the other

Non-Nested = one model is not a direct subset of the other

Two Ways of Conveying Effect Size for Random Effects (Intercepts and Slopes)

- −2ΔLL tests tell us if a random effect is significant, but random effects variances are not likely to have inherent meaning
  - e.g., "I have a significant fixed L1x effect of γ_{10} **= 1.72**, so there is a positive effect on average. I also have a significant L2 random slope variance for L1x of $\tau^2_{U_1}$ **= 0.91**, so groups need their own L1x slopes.
    But how big is a variance of **0.91**, really (i.e., besides different than 0)?"

- We need to convey effect size for random slopes, but pseudo-R^2 is not appropriate because variance has not been explained
  - Fixed effects reduce variance; random effects make new variances (piles)
  - ICC doesn't work for slope variance like it does for intercept variance

- Two ways of conveying effect size for random effects:
  - 95% random effects confidence intervals
  - Indices of random effect reliability

Effect Size via 95% Random Effect CIs

- e.g., "I have a significant fixed L1x effect of γ_{10} **= 1.72**, so there is a positive effect on average. I also have a significant L2 random slope variance for L1x of $\tau^2_{U_1}$ **= 0.91**, so groups need their own L1x slopes. But how big is a variance of **0.91**, really (i.e., besides >0)?"
- **(1) 95% Random Effects Confidence Intervals**
  - Can be calculated for each effect that is random in your model
  - Provide range around the fixed effect within which 95% of YOUR sample is predicted to fall, based on your random effect variance:

$$\text{Random Effect 95\% CI} = \text{fixed effect} \pm \left(1.96 * \sqrt{\text{Random Variance}}\right)$$

$$\text{L1x Slope 95\% CI} = \gamma_{10} \pm \left(1.96 * \sqrt{\tau^2_{U_1}}\right) \rightarrow 1.72 \pm \left(1.96 * \sqrt{0.91}\right) = -0.15 \text{ to } 3.59$$

  - So although L1x has a positive effect on average (its fixed effect), individual group slopes are predicted to range from −0.15 to 3.59 (so some groups are predicted to have a negative L1x slope instead)
  - Is NOT the same as CI for fixed effect using fixed effect SE

Effect Size via Reliability Indices

(2): How reliable is a given level-2 group's random effect?

Intercept Reliability (IR):

$\tau^2_{U_0}$ = random intercept variance

σ^2_e = residual variance

$L1n$ = L1 sample size per L2 unit

$$\mathbf{IR} = \frac{\tau^2_{U_0}}{\tau^2_{U_0} + \frac{\sigma^2_e}{L1n * 1}}$$

Slope Reliability (SR):

$\tau^2_{U_1}$ = random slope variance

σ^2_e = residual variance

$L1n$ = L1 sample size per L2 unit

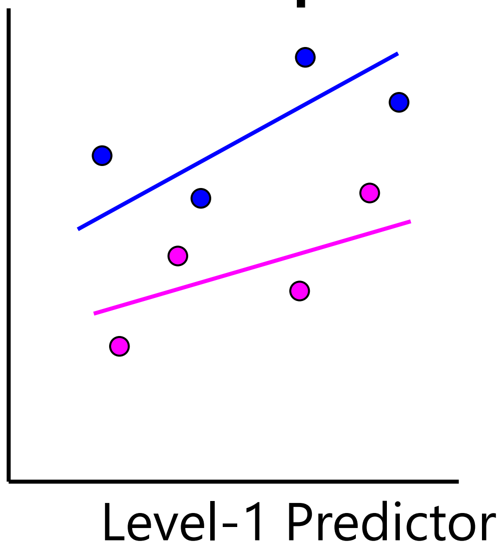
σ^2_{L1} = variance of L1 predictor

$$\mathbf{SR} = \frac{\tau^2_{U_1}}{\tau^2_{U_1} + \frac{\sigma^2_e}{L1n * \sigma^2_{L1}}}$$

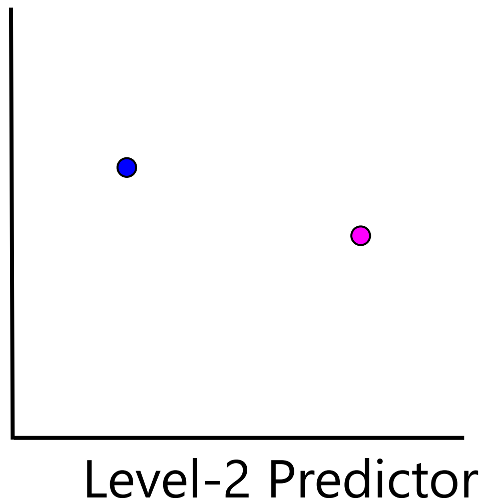
Although these reliability indices are not commonly reported in many fields, they can be very useful for power analyses.

Why Level-2 Predictors Cannot Have Random Effects in Two-Level Models

Random Slopes for L1x



Random Slopes for L2x?



You cannot make a line out of a dot, so level-2 effects cannot vary randomly over groups when groups are independent (two-level model).

Intermediate Summary

- Level-2 predictors refer to Group-Level Variables
  - Can have fixed effects, but not random effects in a two-level model
  - e.g., Does mean school achievement differ b/t rural and urban schools?
- Level-1 predictors refer to Person-Level Variables
  - Can have fixed and/or random effects over groups
  - e.g., Does student achievement differ between boys and girls?
    - Fixed effect: e.g., Is there a gender difference in achievement on average?
    - Random effect: e.g., Does the gender effect differ *randomly* across schools?
  - When a level-1 predictor has both a fixed and random effect, the fixed effect is the average effect across level-2 units (groups)
    - The level-1 fixed effect may differ before vs. after adding a random effect when groups have different numbers of persons (are unbalanced) for this reason
- Random slope variances (and covariances) are tested using $-2\Delta LL$
  - If using REML, to-be-compared models must have same fixed effects

Adding a Cross-Level Interaction to a Grand-MC Level-1 Model: Example

- Level-1 students (p) nested in level-2 schools (g),level-1 x_{pg} = time parents spend reading; new L2 predictor = school in which reading is given "extra emphasis" per external evaluation (0=no, 1=yes)
- **Level-1:** $\textbf{Test}_{pg} = \beta_{0g} + \beta_{1g}(\textbf{Read}_{pg} - C_1) + e_{pg}$
- **Level-2:** $\beta_{0g} = \gamma_{00} + \gamma_{01}(\overline{\textbf{Read}}_g - C_2) + \gamma_{02}(\textbf{Emphasis}_g) + U_{0g}$

  $\beta_{1g} = \gamma_{10} + \gamma_{12}(\textbf{Emphasis}_g) + U_{1g}$

- I skipped γ_{11} because it would have been used for $(\overline{\textbf{Read}}_g - C_2)$
- All good, right? Nope—many researchers may mistakenly think so, but this model is now VERY VERY VERY LIKELY to be broken
  - This is the same error as adding a Grand-MC L1 predictor by itself!

Cross-Level Interactions: The Theory

- **Level-1:** $\textbf{Test}_{pg} = \beta_{0g} + \beta_{1g}(\textbf{Read}_{pg} - C_1) + e_{pg}$
- **Level-2:** $\beta_{0g} = \gamma_{00} + \gamma_{01}(\overline{\textbf{Read}}_{g} - C_2) + \gamma_{02}(\textbf{Emphasis}_{g}) + U_{0g}$

  $\beta_{1g} = \gamma_{10} + \gamma_{12}(\textbf{Emphasis}_{g}) + U_{1g}$

Composite equation:

$$\textbf{Test}_{pg} = \left[\gamma_{00} + \gamma_{01}\left(\overline{\textbf{Read}}_{g} - C_2\right) + \gamma_{02}\left(\textbf{Emphasis}_{g}\right) + U_{0g}\right] + \left[\gamma_{10} + \gamma_{12}\left(\textbf{Emphasis}_{g}\right) + U_{1g}\right]\left(\textbf{Read}_{pg} - C_1\right) + e_{pg}$$

- γ_{12} is a "**cross-level**" interaction: between a L1 and a L2 predictor
  - The purpose of a cross-level interaction is to explain (reduce, account for) the random slope variance of its embedded level-1 predictor
  - **Pseudo-R^2 for random slope variance** can reflect its effect size, so you should **ALWAYS test for L2 random slope variance of the L1 effect first**!
- **Here,** γ_{12} represents the idea that part of the reason why some schools have bigger effects of student (within-school) reading on test scores is because they emphasize reading (if so, γ_{12} should explain U_{1g} variance)

Why this Cross-Level Interaction is Broken

- **Level-1:** $\textbf{Test}_{pg} = \beta_{0g} + \beta_{1g}(\text{Read}_{pg} - C_1) + e_{pg}$
- **Level-2:** $\beta_{0g} = \gamma_{00} + \gamma_{01}(\overline{\text{Read}}_g - C_2) + \gamma_{02}(\text{Emphasis}_g) + U_{0g}$

  $\beta_{1g} = \gamma_{10} + \gamma_{12}(\text{Emphasis}_g) + U_{1g}$

- γ_{00} = fixed intercept: expected test score for a student with reading = C_1 from a school with school mean reading = C_2 and emphasis = no
- γ_{10} = *simple* fixed <u>L1 WG slope</u>: difference in test score per one-unit higher in student (within-school) reading than others in your school *for schools with reading emphasis = no*
- γ_{01} = fixed <u>L2 contextual slope</u>: incremental difference in test score per one-unit higher in <u>school mean</u> reading than other schools (controlling for student reading; not conditional on emphasis)
- $\gamma_{01} + \gamma_{10}$ = fixed <u>L2 BG slope</u>: difference in test score per one-unit higher in <u>school mean reading</u> than other schools (NOT controlling for student reading; not conditional on emphasis)
- γ_{12} = <u>**smushed interaction**</u>: how the L1 WG slope AND how the L2 BG slope each differ between schools without or with reading emphasis

Unsmushing this Cross-Level Interaction

- **Level-1:** $\textbf{Test}_{pg} = \beta_{0g} + \beta_{1g}(\text{Read}_{pg} - C_1) + e_{pg}$
- **Level-2:** $\beta_{0g} = \gamma_{00} + \gamma_{01}(\overline{\text{Read}_g} - C_2) + \gamma_{02}(\text{Emphasis}_g) + \gamma_{03}(\overline{\text{Read}_g} - C_2)(\text{Emphasis}_g) + U_{0g}$

  $\beta_{1g} = \gamma_{10} + \gamma_{12}(\text{Emphasis}_g) + U_{1g}$

- γ_{10} = *simple* fixed L1 WG slope: difference in test score per one-unit higher in student (within-school) reading than others in your school *for schools with reading emphasis = no* (SAME)
- γ_{01} = ***simple*** fixed L2 contextual slope: incremental difference in test score per one-unit higher in school mean reading than other schools for ***schools with emphasis = no*** (controlling for student reading)
- $\gamma_{01} + \gamma_{10}$ = ***simple*** fixed L2 BG slope for ***schools with emphasis = no***
- γ_{12} = how the L1 WG slope differs between schools without or with reading emphasis (NOW UNSMUSHED)
- γ_{03} = how the L2 contextual slope differs between schools without or with reading emphasis (ADDED TO UNSMUSH INTERACTION)
- $\gamma_{12} + \gamma_{03}$ = how the L2 BG slope differs between schools without or with reading emphasis

Unsmushing this Cross-Level Interaction

- **Level-1:** $\textbf{Test}_{pg} = \beta_{0g} + \beta_{1g}(\mathbf{Read}_{pg} - C_1) + e_{pg}$
- **Level-2:** $\beta_{0g} = \gamma_{00} + \gamma_{01}(\overline{\mathbf{Read}}_g - C_2) + \gamma_{02}(\mathbf{Emphasis}_g) + \gamma_{03}(\overline{\mathbf{Read}}_g - C_2)(\mathbf{Emphasis}_g) + U_{0g}$

  $\beta_{1g} = \gamma_{10} + \gamma_{12}(\mathbf{Emphasis}_g) + U_{1g}$

- U_{0g} = level-2 random intercept = deviation between actual and conditional school mean test score predicted for school *g* at $C_1 = 0$
  → for unknown reasons, so is "error"

- U_{1g} = level-2 random slope for effect of Read_{pg} → deviation between the sample mean slope (γ_{10}) and the slope for school *g*
  → Is now <u>conditional: the remaining unknown reasons</u> why some schools have larger student reading effects on scores than others within same school emphasis category
  → Although not shown here, we still have a covariance with U_{0g}

- e_{pg} = level-1 residual = deviation of predicted from student *p*'s real test score

Prerequisites for Cross-Level Interactions?

- What about predicting level-1 effects with no random variance?
  - If the random slope for $(\mathbf{Read}_{pg} - C_1)$ were not significant, can I still test cross-level interactions with it?

$$\beta_{0g} = \gamma_{00} + \gamma_{01}(\overline{\mathbf{Read}}_g - C_2) + \gamma_{02}(\mathbf{Emphasis}_g) + \gamma_{03}(\overline{\mathbf{Read}}_g - C_2)(\mathbf{Emphasis}_g) + U_{0g}$$

$$\beta_{1g} = \gamma_{10} + \gamma_{12}(\mathbf{Emphasis}_g) + U_{1g}$$

Can I still include γ_{12} without U_{1g}?

  - **"NO"**: If a level-1 effect does not vary randomly over groups, then it has "no" variance to predict (so cross-level interactions with that level-1 effect are not necessary); its SE and DDF could be inaccurate SE if $\tau^2_{U_1} \neq 0$
  - **"YES":** Because power to detect random effects is lower than power to detect fixed effects (especially with small L2n), cross-level interactions can still be significant even if there is "no" (≈0) variance to be predicted
  - Saying yes requires new vocabulary…

3 Types of Effects: Fixed, Random, and Systematically (Non-Randomly) Varying

Let's say we have a significant L1 fixed effect of reading. What happens after we test a group*L1x cross-level interaction?

	Non-Significant Group*L1x effect?	Significant Group*L1x effect?
Random L1x slope initially **not** significant	Effect of L1x is **FIXED**	Effect of L1x is **systematically varying**
Random L1x initially sig, **not** sig. after group*L1x	---	Effect of L1x is **systematically varying**
Random L1x initially sig, **still** sig. after group*L1x	Effect of L1x is **RANDOM**	Effect of L1x is **RANDOM**

The effects of level-1 predictors (person-level) can be fixed, random, or systematically varying. The effects of level-2 predictors (group-level) can only be fixed or systematically varying (nothing to be random over…yet).

Variance Accounted For By Level-2 (Group-Level) Predictors

- **Fixed effects of level 2 predictors *by themselves*:**
  - L2 (BG) main effects reduce L2 (BG) random intercept variance
  - L2 (BG) interactions also reduce L2 (BG) random intercept variance
- **Fixed effects of *cross-level interactions* (level 1* level 2):**
  - If the embedded level-1 predictor is random, any cross-level interaction with it will reduce its corresponding level-2 BG **random slope variance**
    - e.g., if *L1x* is random, then any interaction of a group-level predictor with it can reduce the random *L1x* slope variance across groups
  - If the embedded level-1 predictor not random, any cross-level interaction with it will reduce the level-1 WG **residual variance** instead
    - e.g., if *L1x* is fixed, then any interaction of a group-level predictor with it can reduce the L1 (WP) residual variance → Allowing different L1x slopes by group will reduce the level-1 residual variance around those slopes

The Joy of Interactions Involving Level-1 Predictors: By Centering Strategy

- Must consider interactions with both its BG and WG parts:
- Example: Does the effect of employee motivation (x_{pg}) on employee performance interact with type of business (for profit or non-profit; $Type_g$)?

- Group-Mean-Centering:
  - $\mathbf{WGx_{pg}} * \mathbf{Type_g}$ → Does the WG motivation effect differ between business types?
  - $\mathbf{GMx_g} * \mathbf{Type_g}$ → Does the BG motivation effect differ between business types?
    - Moderation of total group motivation effect (not controlling for individual motivation)
    - If forgotten, then $\mathbf{Type_g}$ moderates the motivation effect only at level 1 (WG, not BG)

- Grand-Mean-Centering:
  - $\mathbf{L1x_{pg}} * \mathbf{Type_g}$ → Does the WG motivation effect differ between business types?
  - $\mathbf{GMx_g} * \mathbf{Type_g}$ → Does the *contextual* motivation effect differ b/t business types?
    - Moderation of incremental group motivation effect controlling for employee motivation (moderation of the "boost" in group performance from working with motivated people)
    - If forgotten, then although the level-1 main effect of motivation has been un-smushed via the main effect of $\mathbf{GMx_g}$, the interaction of $\mathbf{L1x_{pg}} * \mathbf{Type_g}$ would still be smushed

Interactions with Level-1 Predictors Example: Employee Motivation (x_{pg}) by Business Type ($Type_g$)

Group-MC: $WGx_{pg} = x_{pg} - GMx_g$

Level-1: $y_{pg} = \beta_{0g} + \beta_{1g}(x_{pg} - GMx_g) + e_{pg}$

Level-2: $\beta_{0g} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{02}(Type_g) + \gamma_{03}(Type_g)(GMx_g) + U_{0g}$

$\beta_{1g} = \gamma_{10} + \gamma_{11}(Type_g)$

Composite: $y_{pg} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg} - GMx_g) + U_{0g} + e_{pg}$
$+ \gamma_{02}(Type_g) + \gamma_{03}(Type_g)(GMx_g) + \gamma_{11}(Type_g)(x_{pg} - GMx_g)$

Grand-MC: $L1x_{pg} = x_{pg}$

Level-1: $y_{pg} = \beta_{0g} + \beta_{1g}(x_{pg}) + e_{pg}$

Level-2: $\beta_{0g} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{02}(Type_g) + \gamma_{03}(Type_g)(GMx_g) + U_{0g}$

$\beta_{1g} = \gamma_{10} + \gamma_{11}(Type_g)$

Composite: $y_{pg} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg}) + U_{0g} + e_{pg}$
$+ \gamma_{02}(Type_g) + \gamma_{03}(Type_g)(GMx_g) + \gamma_{11}(Type_g)(x_{pg})$

Interactions Involving Level-1 Predictors Belong at Both Levels of the Model

On the left below → Group-MC: $WGx_{pg} = x_{pg} - GMx_g$

$y_{pg} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg} - GMx_g) + U_{0g} + e_{pg}$
$+ \gamma_{02}(Type_g) + \gamma_{03}(Type_g)(GMx_g) + \gamma_{11}(Type_g)(x_{pg} - GMx_g)$ ← **As Group-MC**

$y_{pg} = \gamma_{00} + (\gamma_{01} - \gamma_{10})(GMx_g) + \gamma_{10}(x_{pg}) + U_{0g} + e_{pg}$
$+ \gamma_{02}(Type_g) + (\gamma_{03} - \gamma_{11})(Type_g)(GMx_g) + \gamma_{11}(Type_g)(x_{pg})$ ← **As Grand-MC**

On the right below → Grand-MC: $L1x_{pg} = x_{pg}$

$y_{pg} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg}) + U_{0g} + e_{pg}$
$+ \gamma_{02}(Type_g) + \gamma_{03}(Type_g)(GMx_g) + \gamma_{11}(Type_g)(x_{pg})$

After adding an interaction for $Type_g$ with x_{pg} at both levels, then the Group-MC and Grand-MC models are equivalent

---

Intercept: $\gamma_{00} = \gamma_{00}$ **BG Effect:** $\gamma_{01} = \gamma_{01} + \gamma_{10}$ **Contextual:** $\gamma_{01} = \gamma_{01} - \gamma_{10}$

WG Effect: $\gamma_{10} = \gamma_{10}$ **BG*Type Effect:** $\gamma_{03} = \gamma_{03} + \gamma_{11}$ **Contextual*Type:** $\gamma_{03} = \gamma_{03} - \gamma_{11}$

Type Effect: $\gamma_{20} = \gamma_{20}$ **BG*WG or Contextual*WG is the same:** $\gamma_{11} = \gamma_{11}$

Intra-variable Interactions

- Still must consider interactions with both its BG and WG parts!
- Example: Does the effect of employee motivation (x_{pg}) on employee performance interact with business group mean motivation (GMx_g)?

- Group-Mean-Centering:
  - $\mathbf{WGx_{pg}} * \mathbf{GMx_g}$ → Does the WG motivation effect differ by group motivation?
  - $\mathbf{GMx_g} * \mathbf{GMx_g}$ → Does the BG motivation effect differ by group motivation?
    - Moderation of total group motivation effect (not controlling for individual motivation)
    - If forgotten, then $\mathbf{GMx_g}$ moderates the motivation effect only at level 1 (WG, not BG)

- Grand-Mean-Centering:
  - $\mathbf{L1x_{pg}} * \mathbf{GMx_g}$ → Does the WG motivation effect differ by group motivation?
  - $\mathbf{GMx_g} * \mathbf{GMx_g}$ → Does the *contextual* motivation effect differ by group motiv.?
    - Moderation of incremental group motivation effect controlling for employee motivation (moderation of the boost in group performance from working with motivated people)
    - If forgotten, then although the level-1 main effect of motivation has been un-smushed via the main effect of $\mathbf{GMx_g}$, the interaction of $\mathbf{L1x_{pg}} * \mathbf{GMx_g}$ would still be smushed

Intra-variable Interactions Example: Employee Motivation (x_{pg}) by Business Mean Motivation (GMx_g)

Group-MC: $WGx_{pg} = x_{pg} - GMx_g$

Level-1: $y_{pg} = \beta_{0g} + \beta_{1g}(x_{pg} - GMx_g) + e_{pg}$

Level-2: $\beta_{0g} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{02}(GMx_g)(GMx_g) + U_{0g}$

$\beta_{1g} = \gamma_{10} + \gamma_{11}(GMx_g)$

Composite: $y_{pg} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg} - GMx_g) + U_{0g} + e_{pg}$
$+ \gamma_{02}(GMx_g)(GMx_g) + \gamma_{11}(GMx_g)(x_{pg} - GMx_g)$

Grand-MC: $L1x_{pg} = x_{pg}$

Level-1: $y_{pg} = \beta_{0g} + \beta_{1g}(x_{pg}) + e_{pg}$

Level-2: $\beta_{0g} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{02}(GMx_g)(GMx_g) + U_{0g}$

$\beta_{1g} = \gamma_{10} + \gamma_{11}(GMx_g)$

Composite: $y_{pg} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg}) + U_{0g} + e_{pg}$
$+ \gamma_{02}(GMx_g)(GMx_g) + \gamma_{11}(GMx_g)(x_{pg})$

Intra-variable Interactions Example: Employee Motivation (x_{pg}) by Business Mean Motivation (GMx_g)

On the left below → Group-MC: $WGx_{pg} = x_{pg} - GMx_g$

$y_{pg} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg} - GMx_g) + U_{0g} + e_{pg}$
$\quad + \gamma_{02}(GMx_g)(GMx_g) + \gamma_{11}(GMx_g)(x_{pg} - GMx_g)$

$y_{pg} = \gamma_{00} + (\gamma_{01} - \gamma_{10})(GMx_g) + \gamma_{10}(x_{pg}) + U_{0g} + e_{pg}$
$\quad + (\gamma_{02} - \gamma_{11})(GMx_g)(GMx_g) + \gamma_{11}(GMx_g)(x_{pg})$

← **As Group-MC**

← **As Grand-MC**

On the right below → Grand-MC: $L1x_{pg} = x_{pg}$

$y_{pg} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg}) + U_{0g} + e_{pg}$
$\quad + \gamma_{02}(GMx_g)(GMx_g) + \gamma_{11}(GMx_g)(x_{pg})$

After adding an interaction for $Type_g$ with x_{pg} at both levels, then the Group-MC and Grand-MC models are equivalent

Intercept: $\gamma_{00} = \gamma_{00}$ **BG Effect:** $\gamma_{01} = \gamma_{01} + \gamma_{10}$ **Contextual:** $\gamma_{01} = \gamma_{01} - \gamma_{10}$

WG Effect: $\gamma_{10} = \gamma_{10}$ **BG² Effect:** $\gamma_{02} = \gamma_{02} + \gamma_{11}$ **Contextual²:** $\gamma_{02} = \gamma_{02} - \gamma_{11}$

BG*WG or Contextual*WG is the same: $\gamma_{11} = \gamma_{11}$

When Group-MC ≠ Grand-MC: Random Effects of Level-1 Predictors

Group-MC: $WGx_{pg} = x_{pg} - GMx_g$

Level-1: $y_{pg} = \beta_{0g} + \beta_{1g}(x_{pg} - GMx_g) + e_{pg}$

Level-2: $\beta_{0g} = \gamma_{00} + \gamma_{01}(GMx_g) + U_{0g}$

$\beta_{1g} = \gamma_{10} + U_{1g}$

Variance due to GMx_g is removed from the random slope in Group-MC.

$\rightarrow y_{pg} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg} - GMx_g) + U_{0g} + U_{1g}(x_{pg} - GMx_g) + e_{pg}$

Grand-MC: $L1x_{pg} = x_{pg}$

Level-1: $y_{pg} = \beta_{0g} + \beta_{1g}(x_{pg}) + e_{pg}$

Level-2: $\beta_{0g} = \gamma_{00} + \gamma_{01}(GMx_g) + U_{0g}$

$\beta_{1g} = \gamma_{10} + U_{1g}$

Variance due to GMx_g is still part of the random slope in Grand-MC. So these models cannot be made equivalent.

$\rightarrow y_{pg} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg}) + U_{0g} + U_{1g}(x_{pg}) + e_{pg}$

Random Effects of Level-1 Predictors

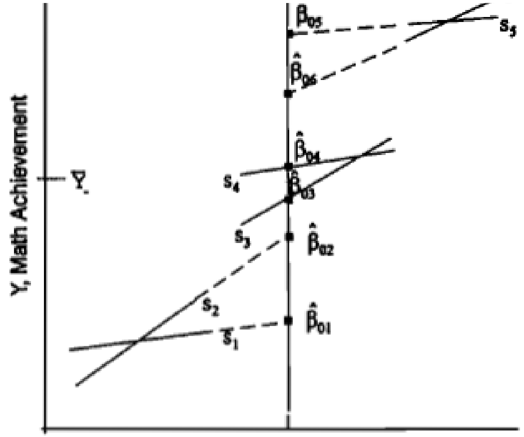
- **Random intercepts** mean different things under each model:
  - **Group-MC** → Group differences at **WGx_{pg} =0** (that every group has)
  - **Grand-MC** → Group differences at **$L1x_{pg}$=0** (that not every group will have)

- **Differential shrinkage of the random intercepts** results from differential reliability of the predicted intercept across models:
  - Group-MC → Won't affect shrinkage of slopes unless highly correlated
  - Grand-MC → Will affect shrinkage of slopes due to forced extrapolation

- As a result, the **random slope variance may be smaller** under Grand-MC than under Group-MC
  - Problem worsens with greater ICC of level-1 predictor (more extrapolation)
  - Anecdotal example was presented in Raudenbush & Bryk (2002; chapter 5)

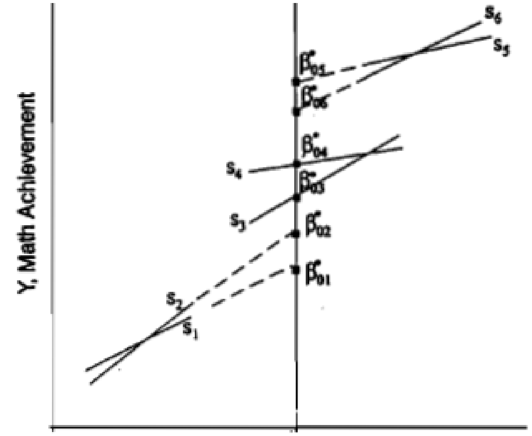
Bias in Random Slope Variance

OLS Per-Group Estimates



Level-1 X

EB Shrunken Estimates



Level-1 X

Top right: Intercepts and slopes are homogenized in Grand-MC because of intercept extrapolation

Bottom: Downwardly-biased random slope variance in Grand-MC relative to Group-MC

Unconditional Results	*Conditional Results*
Group-MC	
$\hat{\mathbf{T}} = \begin{bmatrix} 8.68 & 0.05 \\ 0.05 & 0.68 \end{bmatrix}$	$\hat{\mathbf{T}} = \begin{bmatrix} 2.38 & 0.19 \\ 0.19 & \boxed{0.15} \end{bmatrix}$
$\hat{\sigma}^2 = 36.70$	$\hat{\sigma}^2 = 36.70$
Grand-MC	
$\hat{\mathbf{T}} = \begin{bmatrix} 4.83 & -0.15 \\ -0.15 & 0.42 \end{bmatrix}$	$\hat{\mathbf{T}} = \begin{bmatrix} 2.41 & 0.19 \\ 0.19 & \boxed{0.06} \end{bmatrix}$
$\hat{\sigma}^2 = 36.83$	$\hat{\sigma}^2 = 36.74$

Intermediate Summary

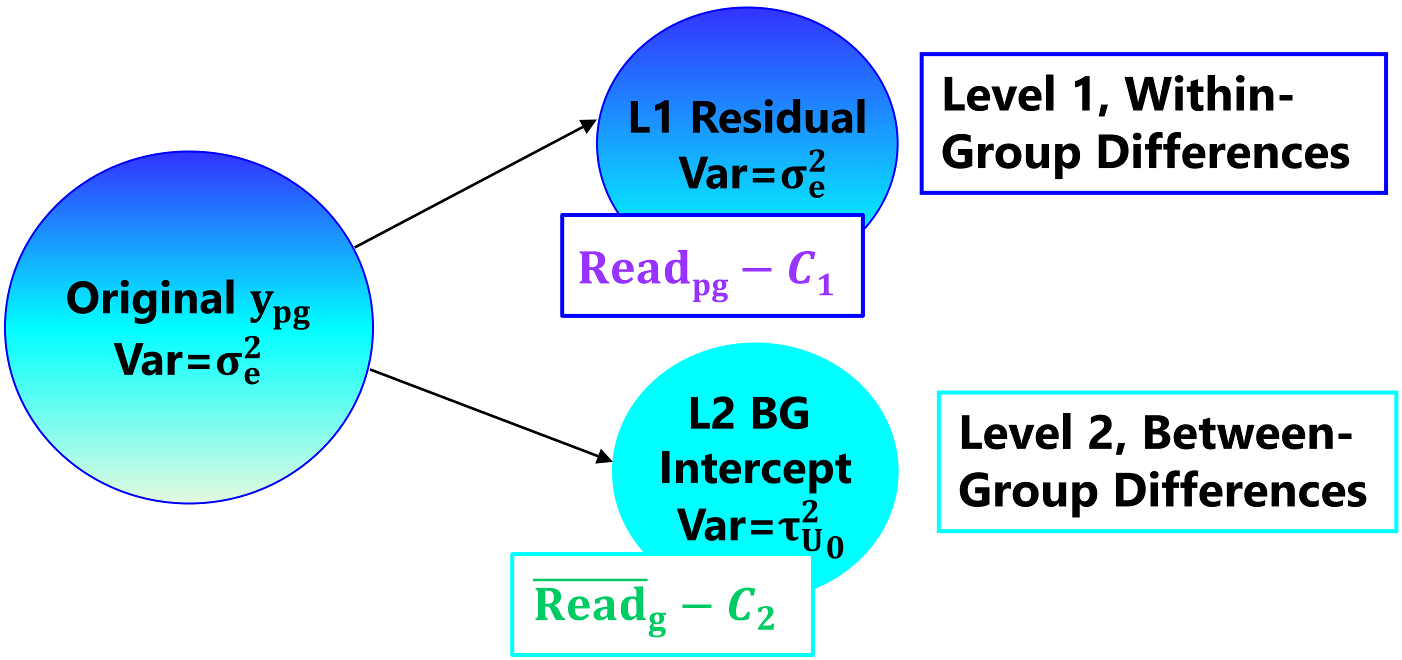
- Each variance component (pile of variance) is either:
  - Unconditional: no predictors of that level-specific term (all the variance)
  - Conditional: variance leftover after predictors of that level-specific term

- L2 = Between-Group, L1 = Within-Group (between-person)
  - Level-2 predictors are group variables: can have fixed or systematically varying effects (but not random effects in two-level models)
  - Level-1 predictors are person variables: can have fixed, random, or systematically varying effects (but check the L1 random slope first!)

- No smushing main effects or interactions of level-1 predictors:
  - Group-MC at Level 1: Get L1=WG and L2=BG effects directly
  - Grand-MC at Level 1: Get L1=WG and L2=contextual effects directly
    - As long as some representation of the L1 effect is included in L2; otherwise, the L1 effect (and any interactions thereof) will be smushed

How MLM “Handles” Dependency

- The purpose of MLM is to “address” or “handle” correlated (dependent) data (such as the dependency of persons from the same group)
- How does it do so? By forming a new random effect variance component (or “pile” of variance) for each source of dependency
- Here, a **random intercept only model**:



Random Intercept Only Models Imply…

- **Groups differ from each other systematically in only ONE way**—in intercept (**U_{0g}**), which implies **ONE kind of BG variance**, which translates to **ONE source of group dependency** (only one reason for covariance and correlation across persons from the same group)
- If this is true, after controlling for L2 BG intercept differences (by estimating the variance of U_{0g} as $\boldsymbol{\tau^2_{U0}}$ in the **G** matrix), the L1 **e_{pg} residuals** (whose variance and covariance are estimated in the **R** matrix) should be **uncorrelated with equal variance across persons**, as shown below for an example of four persons (# of rows and columns in **R**) from a group:

Level-2 **G** matrix: RANDOM (TYPE=UN)

$$\begin{bmatrix} \tau^2_{U_0} \end{bmatrix}$$

Level-1 **R** matrix: REPEATED TYPE=VC

$$\begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

G and **R** matrices create a total **V** matrix with Compound Symmetry

$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

Random Intercept Only Model
(5 total parameters: effect of L1 x_{pg} is FIXED only)

- Example Model:

Level-1: $y_{pg} = \beta_{0g} + \beta_{1g}(x_{pg}) + e_{pg}$

Level-2: $\beta_{0g} = \gamma_{00} + \gamma_{01}(GMx_g) + U_{0g}$

$\beta_{1g} = \gamma_{10}$

Composite Model: $\gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg}) + U_{0g} + e_{pg}$

- This model can be written in matrix form, such as shown below for each group:

$$Y_g = X_g * \gamma + Z_g * U_g + E_g$$

Btw—this equation is where the terms "**columns in X**" and "**columns in Z**" on the SAS MIXED output come from

Scalar and Matrix "Mixed Model" Equations: Here, per Group of n = 4 Persons

Random Int Only: $\gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg}) + U_{0g} + e_{pg}$

$$\mathbf{Y}_g = \mathbf{X}_g * \boldsymbol{\gamma} + \mathbf{Z}_g * \mathbf{U}_g + \mathbf{E}_g$$

$$\begin{bmatrix} y_{1g} \\ y_{2g} \\ y_{3g} \\ y_{4g} \end{bmatrix} = \begin{bmatrix} 1 & GMx_g & x_{1g} \\ 1 & GMx_g & x_{2g} \\ 1 & GMx_g & x_{3g} \\ 1 & GMx_g & x_{4g} \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} U_{0g} \end{bmatrix} + \begin{bmatrix} e_{1g} \\ e_{2g} \\ e_{3g} \\ e_{4g} \end{bmatrix}$$

$$\begin{bmatrix} y_{1g} \\ y_{2g} \\ y_{3g} \\ y_{4g} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{1g}) \\ \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{2g}) \\ \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{3g}) \\ \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{4g}) \end{bmatrix} + \begin{bmatrix} U_{0g} \\ U_{0g} \\ U_{0g} \\ U_{0g} \end{bmatrix} + \begin{bmatrix} e_{1g} \\ e_{2g} \\ e_{3g} \\ e_{4g} \end{bmatrix}$$

$$\begin{bmatrix} y_{1g} \\ y_{2g} \\ y_{3g} \\ y_{4g} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{1g}) + U_{0g} + e_{1g} \\ \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{2g}) + U_{0g} + e_{2g} \\ \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{3g}) + U_{0g} + e_{3g} \\ \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{4g}) + U_{0g} + e_{4g} \end{bmatrix}$$

$\mathbf{X}_g$ = n x k values of **predictors with fixed effects**, so can differ per group (k = 3 here)

$\boldsymbol{\gamma}$ = k x 1 estimated **fixed effects** → same for all groups

$\mathbf{Z}_g$ = n x u values of **predictors with random effects**, so can differ per group (u = 1 here)

$\mathbf{U}_g$ = u x 1 estimated group-specific **random effects**

$\mathbf{E}_g$ = n x n person-specific within-group residuals

Same Random Intercept Only Model: Predicted Total Variance-Covariance V Matrix per Group

$$\mathbf{V_g} = \mathbf{Z_g} * \mathbf{G_g} * \mathbf{Z_g^T} + \mathbf{R_g}$$

$$\mathbf{V_g} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}\tau^2_{U_0}\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}\sigma^2_e & 0 & 0 & 0\\ 0 & \sigma^2_e & 0 & 0\\ 0 & 0 & \sigma^2_e & 0\\ 0 & 0 & 0 & \sigma^2_e\end{bmatrix}$$

$$\mathbf{V_g} = \begin{bmatrix}\tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e\end{bmatrix}$$

$\mathbf{Z_g}$ = *n* x *u* values of **predictors with random effects**, so can differ per group (*u* = 1: intercept)

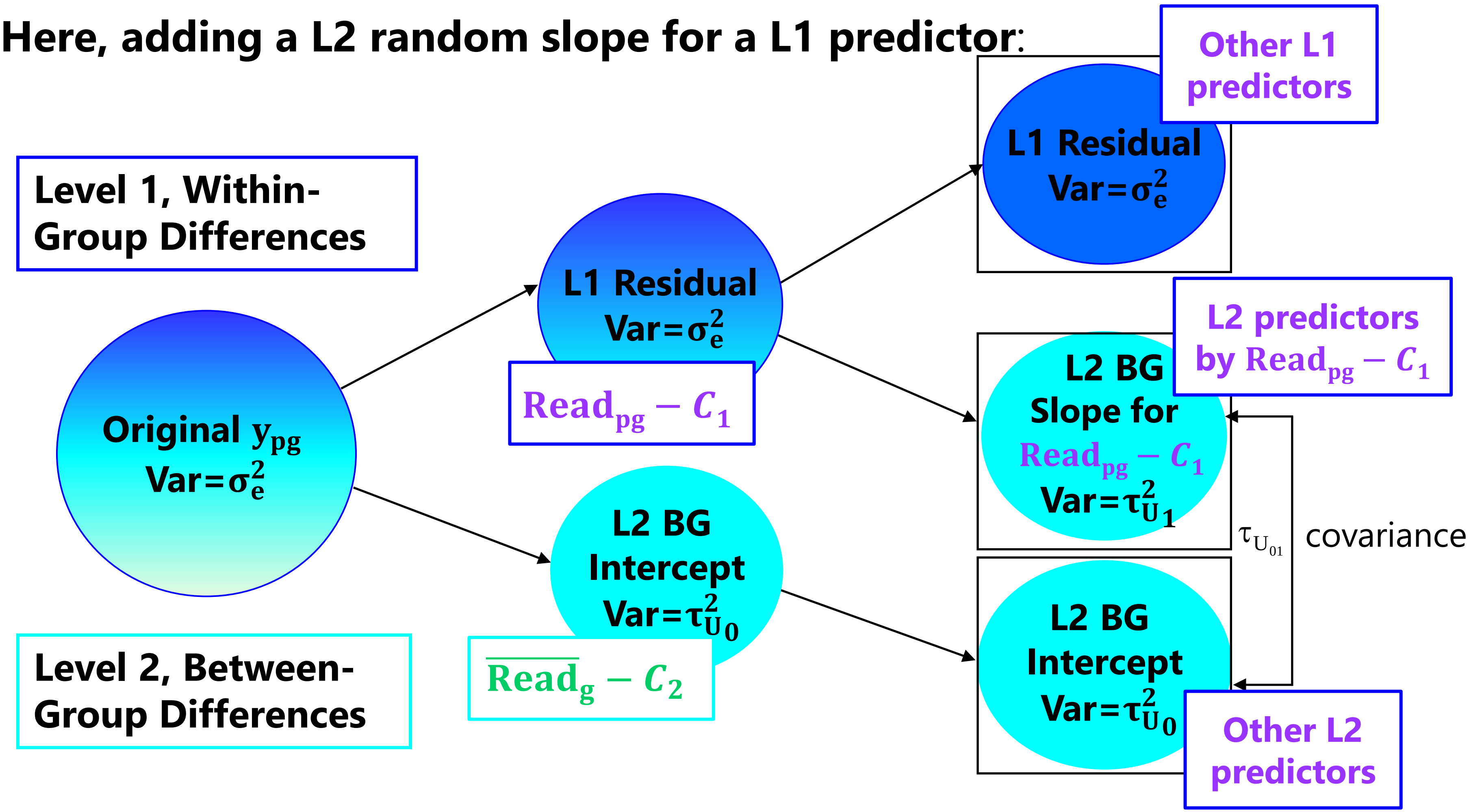
$\mathbf{Z_g}^T$ = *u* x *n* values of predictors with random effects (just $\mathbf{Z_g}$ transposed)

$\mathbf{G_g}$ = *u* x *u* estimated **random effects variances and covariances**, so will be the same for all groups (just $\boldsymbol{\tau^2_{U0}}$ = intercept variance here)

$\mathbf{R_g}$ = *n* x *n* **time-specific residual variances and covariances**, so will be same for all groups (here, just $\boldsymbol{\sigma^2_e}$ on the diagonal)

How MLM "Handles" Dependency

- How does MLM "handle" dependency? By forming a new random effect variance component (or "pile" of variance) for each source of dependency
- **Here, adding a L2 random slope for a L1 predictor**:



Adding a Random Slope Implies:

- **Groups differ from each other systematically in TWO ways**—in intercept (**U_{0g}**) and slope (**U_{1g}**), which implies **TWO kinds of BG variance**, which translates to **TWO sources of group dependency** (two reasons for covariance or correlation across persons from the same group)
- If so, after controlling for both BG intercept and slope differences (via the $\tau^2_{U_0}$ and $\tau^2_{U_1}$ variances in the L2 **G** matrix), the e_{pg} **residuals** (whose variance and covariance are estimated in the L1 **R** matrix) should be **uncorrelated with equal variance across persons**, as shown below for an example of four persons (# of rows and columns in **R**) from a group:

Level-2 **G** matrix: RANDOM TYPE=UN

$$\begin{bmatrix} \tau^2_{U_0} & \tau_{U_{10}} \\ \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix}$$

Level-1 **R** matrix: REPEATED TYPE=VC

$$\begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

G and **R** combine to create a total **V** matrix whose per-group structure depends on the specific L1x values each person has and group-specific size (very flexible)

Unconditional Random Slope Model
(7 total parameters: effect of L1 x_{pg} is RANDOM)

- Example Model:

Level-1: $y_{pg} = \beta_{0g} + \beta_{1g}(x_{pg}) + e_{pg}$

Level-2: $\beta_{0g} = \gamma_{00} + \gamma_{01}(GMx_g) + U_{0g}$

$\beta_{1g} = \gamma_{10} + U_{1g}$

Composite Model: $\gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg})$
$+ U_{0g} + U_{1g}(x_{pg}) + e_{pg}$

- This model can be written in matrix form, such as shown below for each group:

$$\mathbf{Y}_g = \mathbf{X}_g * \gamma + \mathbf{Z}_g * \mathbf{U}_g + \mathbf{E}_g$$

Btw—this equation is where the terms "**columns in X**" and "**columns in Z**" on the SAS MIXED output come from

"Mixed Model" Equations with Random Slope: Here, per Group of n = 4 Persons

Model: $\gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg}) + U_{0g} + U_{1g}(x_{pg}) + e_{pg}$

$$Y_g = X_g * \gamma + Z_g * U_g + E_g$$

$$\begin{bmatrix} y_{1g} \\ y_{2g} \\ y_{3g} \\ y_{4g} \end{bmatrix} = \begin{bmatrix} 1 & GMx_g & x_{1g} \\ 1 & GMx_g & x_{2g} \\ 1 & GMx_g & x_{3g} \\ 1 & GMx_g & x_{4g} \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} 1 & x_{1g} \\ 1 & x_{2g} \\ 1 & x_{3g} \\ 1 & x_{4g} \end{bmatrix} \begin{bmatrix} U_{0g} \\ U_{1g} \end{bmatrix} + \begin{bmatrix} e_{1g} \\ e_{2g} \\ e_{3g} \\ e_{4g} \end{bmatrix}$$

$$\begin{bmatrix} y_{1g} \\ y_{2g} \\ y_{3g} \\ y_{4g} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{1g}) \\ \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{2g}) \\ \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{3g}) \\ \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{4g}) \end{bmatrix} + \begin{bmatrix} U_{0g} + U_{1g}(x_{1g}) \\ U_{0g} + U_{1g}(x_{2g}) \\ U_{0g} + U_{1g}(x_{3g}) \\ U_{0g} + U_{1g}(x_{4g}) \end{bmatrix} + \begin{bmatrix} e_{1g} \\ e_{2g} \\ e_{3g} \\ e_{4g} \end{bmatrix}$$

$$\begin{bmatrix} y_{1g} \\ y_{2g} \\ y_{3g} \\ y_{4g} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{1g}) + U_{0g} + U_{1g}(x_{1g}) + e_{1g} \\ \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{2g}) + U_{0g} + U_{1g}(x_{2g}) + e_{2g} \\ \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{3g}) + U_{0g} + U_{1g}(x_{3g}) + e_{3g} \\ \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{4g}) + U_{0g} + U_{1g}(x_{4g}) + e_{4g} \end{bmatrix}$$

$\mathbf{X_g}$ = *n* x *k* values of **predictors with fixed effects**, so can differ per group (*k* = 3 here)

$\boldsymbol{\gamma}$ = *k* x 1 estimated **fixed effects** → same for all groups

$\mathbf{Z_g}$ = *n* x *u* values of **predictors with random effects**, so can differ per group (*u* = 2 here)

$\mathbf{U_g}$ = *u* x 1 estimated group-specific **random effects**

$\mathbf{E_g}$ = *n* x *n* person-specific residuals

Same Random Slope Model: Predicted Total Variance-Covariance V Matrix per Group

$$\mathbf{V}_g \quad = \quad \mathbf{Z}_g * \mathbf{G}_g \quad * \quad \mathbf{Z}_g^T \quad + \quad \mathbf{R}_g$$

$$\mathbf{V}_g = \begin{bmatrix} 1 & x_{1g} \\ 1 & x_{2g} \\ 1 & x_{3g} \\ 1 & x_{4g} \end{bmatrix} \begin{bmatrix} \tau_{U_0}^2 & \tau_{U_{01}} \\ \tau_{U_{01}} & \tau_{U_1}^2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_{1g} & x_{2g} & x_{3g} & x_{4g} \end{bmatrix} + \begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$$

$\mathbf{V}_g$ matrix = complicated, but summarized below

$\tau_{U_{01}}$

$$\mathbf{V}_g \text{ matrix: Outcome Variance Given}[x]$$

$$= \tau_{U_0}^2 + \left[\left(x^2\right)\tau_{U_1}^2\right] + \left[2(x)\tau_{U_{01}}\right] + \sigma_e^2$$

$$\mathbf{V}_g \text{ matrix: Outcome Covariance Given}[x_A, y_B]$$

$$= \tau_{U_0}^2 + \left[(A+B)\tau_{U_{01}}\right] + \left[(AB)\tau_{U_1}^2\right]$$

$\mathbf{Z}_g$ = *n* x *u* values of **predictors with random effects**, so can differ per group (*u* = 2: int, slope)

$\mathbf{Z}_g^T$ = *u* x *n* values of predictors with random effects (just **$\mathbf{Z}_g$** transposed)

$\mathbf{G}_g$ = *u* x *u* estimated **random effects variances and covariances**, so will be the same for all groups ($\boldsymbol{\tau}_{U_0}^2$, $\boldsymbol{\tau}_{U_1}^2$ and cov $\boldsymbol{\tau}_{U_{01}}$)

$\mathbf{R}_g$ = *n* x *n* **time-specific residual variances and covariances**, so will be same for all groups (here, just $\boldsymbol{\sigma}_e^2$ on the diagonal)

Building V across Groups: Same Random Slope Model

- **V** for two groups of equal size n = 4:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R}$$

$$\mathbf{V} = \begin{bmatrix} 1 & 0.0 & 0 & 0 \\ 1 & 1.0 & 0 & 0 \\ 1 & 2.0 & 0 & 0 \\ 1 & 3.0 & 0 & 0 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1.4 \\ 0 & 0 & 1 & 2.3 \\ 0 & 0 & 1 & 3.5 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} & 0 & 0 \\ \tau_{U_{01}} & \tau^2_{U_1} & 0 & 0 \\ 0 & 0 & \tau^2_{U_0} & \tau_{U_{01}} \\ 0 & 0 & \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0.0 & 1.0 & 2.0 & 3.0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.2 & 1.4 & 2.3 & 3.5 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

- The giant combined **V** matrix across groups is how the multilevel or mixed model is actually estimated in SAS
- It has a "**block diagonal**" structure → predictions are given for each group, but 0's are given for the elements that describe relationships across groups (because groups are supposed to be independent here!)

Building **V** across Groups: Same Random Slope Model

- **V** for two groups also with **different *n*** per group:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R}$$

$$\mathbf{V} = \begin{bmatrix} 1 & 0.0 & 0 & 0 \\ 1 & 1.0 & 0 & 0 \\ 1 & 2.0 & 0 & 0 \\ 1 & 3.0 & 0 & 0 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1.4 \\ 0 & 0 & 1 & 3.5 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} & 0 & 0 \\ \tau_{U_{01}} & \tau^2_{U_1} & 0 & 0 \\ 0 & 0 & \tau^2_{U_0} & \tau_{U_{01}} \\ 0 & 0 & \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0.0 & 1.0 & 2.0 & 3.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.2 & 1.4 & 3.5 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

- Take Home message: Partitioning variance into piles…
  - **Level 2 = BG →** **G** matrix of random effects variances/covariances
  - **Level 1 = WG →** **R** matrix of residual variances/covariances
  - **G** and **R** combine via **Z** to create **V** matrix of total variances/covariances
  - These flexible options allow the outcome variances and covariances to vary in a predictor-dependent way to better match the actual data

Two Sides of Any Model: Estimation

- **Fixed Effects in the Model for the Means:**
  - How the expected outcome for a given observation varies as a function of values on *known* predictor variables
  - Fixed effects predict the Y values *are not parameters that are solved for iteratively in maximum likelihood estimation for general MLMs*

- **Random Effects in the Model for the Variance:**
  - How model residuals are related across observations (dependency across persons, groups, time, etc) – *unknown* things due to sampling
  - Random effects variances and covariances are a mechanism by which complex patterns of variance and covariance among the Y residuals can be predicted (not the Y values, but their dispersion)
  - Anything besides level-1 residual variance $\boldsymbol{\sigma_e^2}$ must be solved for iteratively—this increases the dimensionality of estimation process
  - Estimation utilizes the predicted **V** matrix for each group
  - In what follows, **V** will be based on previous random slope model

End Goals of Maximum Likelihood Estimation

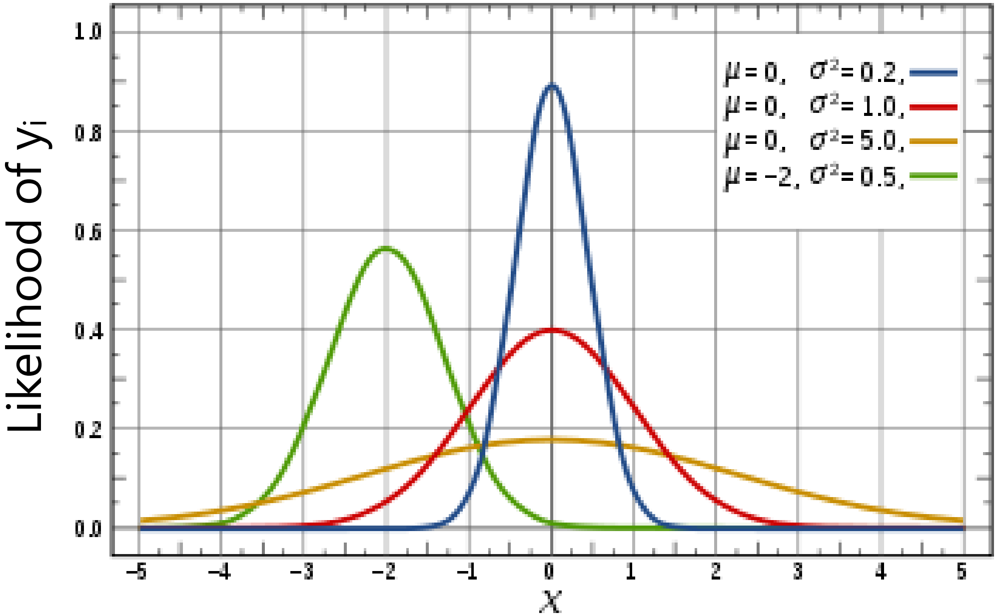
1. Obtain "most likely" values for each unknown model parameter (random effects variances and covariances, residual variances and covariances, which then are used to calculate the fixed effects) → **the estimates**

2. Obtain an index as to how likely each parameter value actually is (i.e., "really likely" or pretty much just a guess?) → **the standard error (SE) of the estimates**

3. Obtain an index as to how well the model we've specified actually describes the data → **the model fit indices**

How does all this happen? The magic of multivariate normal…(but let's start with univariate normal first)

Univariate Normal PDF



- This function tells us how **likely** any value of y_i is given two pieces of info:
  - predicted value $\hat{y}_i$
  - residual variance σ_e^2

- Example: regression

$$y_i = \beta_0 + \beta_1 X_i + e_i$$

$$\hat{y}_i = \beta_0 + \beta_1 X_i$$

$$e_i = y_i - \hat{y}_i \qquad \sigma_e^2 = \frac{\sum_{i=1}^{N} e_i^2}{N-2}$$

Univariate Normal PDF (two ways):

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma_e^2}} * \exp\left[-\frac{1}{2} * \frac{\left(y_i - \hat{y}_i\right)^2}{\sigma_e^2}\right]$$

$$f(y_i) = \left(2\pi\sigma_e^2\right)^{-1/2} * \exp\left[-\frac{1}{2} * \left(y_i - \hat{y}_i\right)\left(\sigma_e^2\right)^{-1}\left(y_i - \hat{y}_i\right)\right]$$

Multivariate Normal for Y_g (height for all *n* outcomes for group *g*)

$$\text{Univariate Normal PDF: } f(y_i) = \left(2\pi\sigma_e^2\right)^{-1/2} * \exp\left[-\frac{1}{2} * \left(y_i - \hat{y}_i\right)\left(\sigma_e^2\right)^{-1}\left(y_i - \hat{y}_i\right)\right]$$

$$\text{Multivariate Normal PDF: } f(\mathbf{Y}_g) = \left(2\pi\right)^{-n/2} * \left|\mathbf{V}_g\right|^{-1/2} * \exp\left[-\frac{1}{2} * \left(\mathbf{Y}_g - \mathbf{X}_g\gamma\right)^T \left(\mathbf{V}_g\right)^{-1}\left(\mathbf{Y}_g - \mathbf{X}_g\gamma\right)\right]$$

- In our random slope model, three are three fixed effects (in γ) that predict the $\mathbf{Y}_g$ outcome values: intercept γ_{00}, L2 slope γ_{01}, and L1 slope γ_{10}
- The model also gives us $\mathbf{V}_g$ → the model-predicted total variance and covariance matrix across the occasions, taking into account the time values
- Uses $|\mathbf{V}_g|$ = determinant of $\mathbf{V}_g$= summary of *non-redundant* info
  - Reflects sum of variances across occasions controlling for covariances
- $(\mathbf{V}_g)^{-1}$ → matrix inverse → like dividing (so can't be 0 or negative)
  - $(\mathbf{V}_g)^{-1}$ must be "positive definite", which in practice means no 0 random variances or covariances that cause out-of-bound correlations between random effects
  - Otherwise, program uses "generalized inverse" → questionable results

Now Try Some Possible Answers...

(e.g., for the 4 $\mathbf{V}$ parameters in this random slope model example)

- Plug $\mathbf{V}_g$ predictions into log-likelihood function, sum over groups:

$$L = \prod_{g=1}^{N} \left\{ (2\pi)^{-n/2} * \left|\mathbf{V}_g\right|^{-1/2} * \exp\left[-\frac{1}{2} \left(\mathbf{Y}_g - \mathbf{X}_g \boldsymbol{\gamma}\right)^T \left(\mathbf{V}_g\right)^{-1} \left(\mathbf{Y}_g - \mathbf{X}_g \boldsymbol{\gamma}\right) \right] \right\}$$

$$LL = \sum_{g=1}^{N} \left\{ \left[-\frac{n}{2} \log(2\pi) \right] + \left[-\frac{1}{2} \log\left|\mathbf{V}_g\right| \right] + \left[-\frac{1}{2} \left(\mathbf{Y}_g - \mathbf{X}_g \boldsymbol{\gamma}\right)^T \left(\mathbf{V}_g\right)^{-1} \left(\mathbf{Y}_g - \mathbf{X}_g \boldsymbol{\gamma}\right) \right] \right\}$$

- Try one set of possible parameter values for $\mathbf{V}_g$, compute LL
- Try another possible set for $\mathbf{V}_g$, compute LL….
  - Different algorithms are used to decide which values to try given that each parameter has its own distribution → like an uncharted mountain
  - Calculus helps the program scale this multidimensional mountain
    - At the top, all first partial derivatives (linear slopes at that point) ≈ 0
    - Positive first partial derivative? Too *low*, try again. Negative? Too *high*, try again.
    - Matrix of partial first derivatives = "score function" = "gradient" (as given in SAS GLIMMIX or NLMIXED output for generalized or truly nonlinear effects models)

End Goals 1 and 2: Model Estimates and SEs

- Process terminates (the model "converges") when the next set of tried values for $\mathbf{V_g}$ don't improve the LL very much…
  - e.g., SAS default convergence criteria = .00000001
  - Those are the values for the parameters that, relative to the other possible values tried, are "most likely" → the variance estimates

- But we need to know how trustworthy those estimates are…
  - Precision is indexed by the steepness of the multidimensional mountain, where steepness → more negative partial second derivatives
  - Matrix of partial second derivatives = "Hessian matrix"
  - Hessian matrix * -1 = "information matrix"
  - So steeper function = more information = more precision = smaller SE

$$\text{Each parameter SE} = \frac{1}{\sqrt{\text{information}}}$$

What about the Fixed Effects?

- Likelihood mountain does NOT include fixed effects as additional search dimensions (only variances and covariances that make $\mathbf{V}_g$)
- **Fixed effects are determined** given the parameters for $\mathbf{V}_g$:

$$\gamma = \left\{\sum_{g=1}^{N}\left(\mathbf{X}_g^T \mathbf{V}_g^{-1} \mathbf{X}_g\right)\right\}^{-1} \sum_{g=1}^{N}\left(\mathbf{X}_g^T \mathbf{V}_g^{-1} \mathbf{Y}_g\right), \qquad \text{Cov}(\gamma) = \left\{\sum_{g=1}^{N}\left(\mathbf{X}_g^T \mathbf{V}_g^{-1} \mathbf{X}_g\right)\right\}^{-1}$$

All we need is $\mathbf{V}_g$ and the data: $\mathbf{X}_g$, $\mathbf{Y}_g$

γ = fixed effect estimates

Cov(γ) = γ sampling variance (SQRT of diagonal = SE)

- This is actually what happens in regular regression (GLM), too:

$$\text{GLM matrix solution: } \boldsymbol{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\mathbf{Y}\right), \qquad \text{Cov}(\boldsymbol{\beta}) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\sigma_e^2$$

$$\text{GLM scalar solution: } \beta = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}, \qquad \text{Cov}(\beta) = \frac{\sigma_e^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

- **Implication: fixed effects don't cause estimation problems**… (at least in general linear mixed models with normal residuals)

What about ML vs. REML?

- **REML** estimates of random effects variances and covariances are **unbiased** because they account for the uncertainty that results from simultaneously also estimating fixed effects (whereas ML estimates do not, so they are too small)

- What does this mean? Remember "population" vs. "sample" formulas for computing variance?

$$\text{Population: } \sigma_e^2 = \frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N} \qquad \text{Sample: } \sigma_e^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}$$

  - $N - 1$ is used because the mean had to be estimated from the data (i.e., the mean is the fixed intercept)…

- Same idea: ML estimates of random effects variances will be downwardly biased by a factor of $(N - k) / N$, where N = # persons and k = #fixed effects… it just looks way more complicated

What about ML vs. REML?

ML: $$\text{LL} = \left[-\frac{T-0}{2}\log(2\pi)\right] + \left[-\frac{1}{2}\sum_{g=1}^{N}\log\left|\mathbf{V}_g\right|\right] + \left[-\frac{1}{2}\sum_{g=1}^{N}\left(\mathbf{Y}_g - \mathbf{X}_g\boldsymbol{\gamma}\right)^T \mathbf{V}_g^{-1}\left(\mathbf{Y}_g - \mathbf{X}_g\boldsymbol{\gamma}\right)\right]$$

REML: $$\text{LL} = \left[-\frac{T-k}{2}\log(2\pi)\right] + \left[-\frac{1}{2}\sum_{g=1}^{N}\log\left|\mathbf{V}_g\right|\right] + \left[-\frac{1}{2}\sum_{g=1}^{N}\left(\mathbf{Y}_g - \mathbf{X}_g\boldsymbol{\gamma}\right)^T \mathbf{V}_g^{-1}\left(\mathbf{Y}_g - \mathbf{X}_g\boldsymbol{\gamma}\right)\right]$$

$$+ \left[-\frac{1}{2}\log\left|\sum_{g=1}^{N}\mathbf{X}_g^T\mathbf{V}_g^{-1}\mathbf{X}_g\right|\right]$$

where: $$\left[-\frac{1}{2}\log\left|\sum_{g=1}^{N}\mathbf{X}_g^T\mathbf{V}_g^{-1}\mathbf{X}_g\right|\right] = \left[\frac{1}{2}\log\left|\left(\sum_{g=1}^{N}\mathbf{X}_g^T\mathbf{V}_g^{-1}\mathbf{X}_g\right)^{-1}\right|\right] = \left[\frac{1}{2}\log\left|\text{Cov}(\boldsymbol{\gamma})\right|\right]$$

- Extra part in REML is the sampling variance of the fixed effects… it is added back in to account for uncertainty in estimating fixed effects
- REML maximizes the likelihood of the residuals specifically, so models with different fixed effects are not on the same scale and are not comparable
  - This is why you can't do −2ΔLL tests in REML when the models to be compared have different fixed effects → the model residuals will be defined differently

End Goal #3: How well do the model predictions match the data?

- End up with ML or REML LL from predicting $\mathbf{V_g}$ → so how good is it?
- Absolute model fit assessment is only possible when the $\mathbf{V_g}$ matrix is organized the same for all L2 units and there are no random slopes
  - Items are usually fixed, so can get absolute fit in CFA and SEM → χ^2 test is based on match between actual and predicted data matrix
  - No absolute fit provided in MLM for unbalanced occasions or groups (or in SEM when using random slopes or T-scores for unbalanced time)
    - Can compute absolute fit when the saturated means, unstructured variance model is estimable in ML → is -2ΔLL versus "perfect" model for balanced time

- Relative model fit is given as **−2LL** in SAS, in which smaller is better
  - -2* needed to conduct "likelihood ratio" or "deviance difference" tests
  - Also information criteria:
    - **AIC**: −2LL + 2*(#parms)
    - **BIC**: −2LL + log(N)*(#parms)
    - #parms = all parameters in ML; #parms = variance model parms only in REML

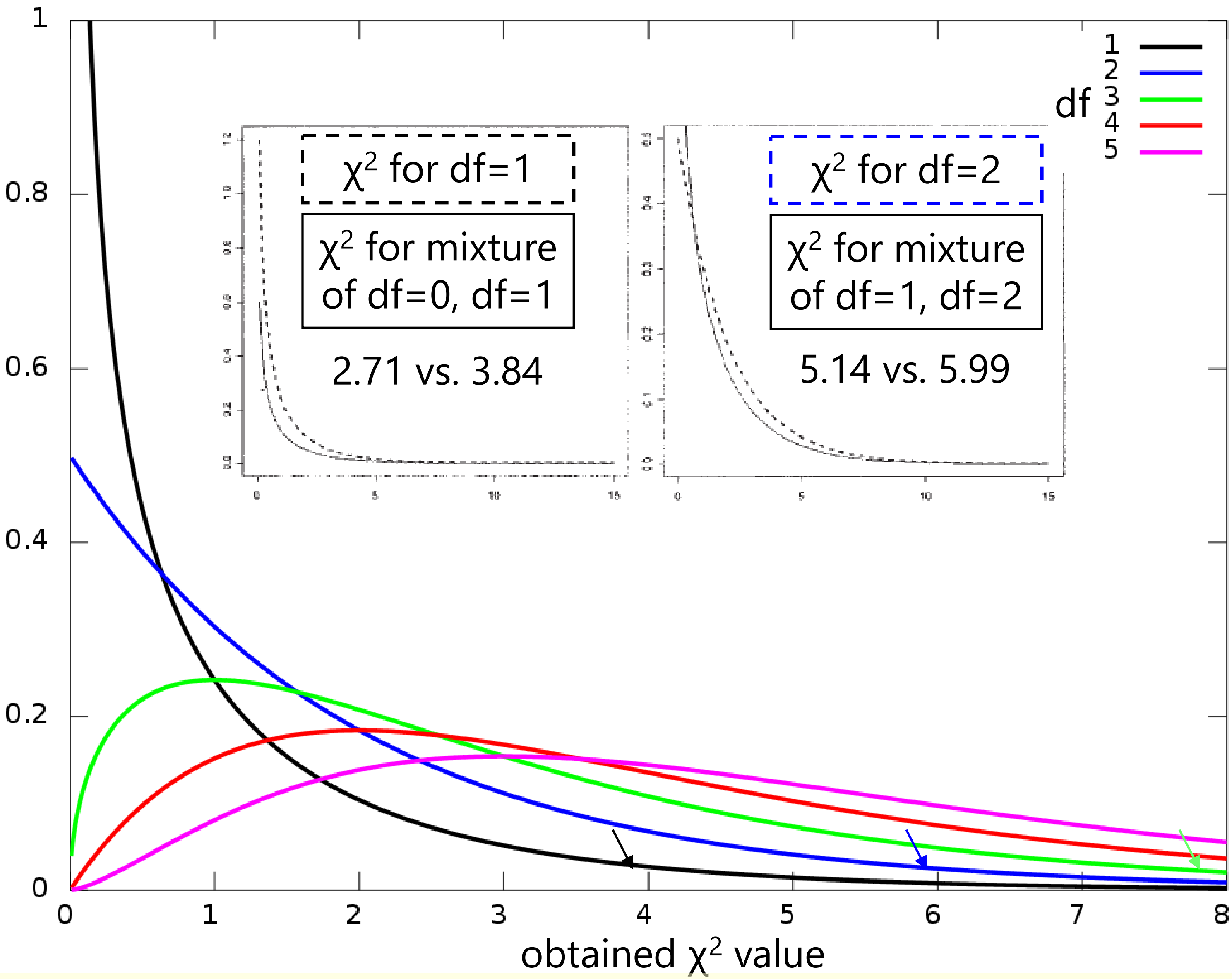
What about testing variances > 0?

- $-2\Delta LL$ between two nested models is distributed as χ^2 only when added parameters do not have a boundary (like 0 or 1)
  - Is ok for fixed effects using ML (could be any positive or negative value)
  - NOT ok for ML or REML tests of random variances (must be > 0)
  - Ok for ML or REML tests of heterogeneous variances and covariances (because extra parameters can be phrased as unbounded deviations)

- When testing addition of parameters that have a boundary, $-2\Delta LL$ will follow a **mixture** of χ^2 distributions instead
  - e.g., when adding random intercept variance (test > 0?)
    - When estimated as positive, will follow χ^2 with df=1
    - When estimated as negative… can't happen, will follow χ^2 with df=0
  - End result: **$-2\Delta LL$ will be too conservative in boundary cases**

χ^2 Distributions

small pictures from Stoel et al., 2006



Critical Values for 50:50 Mixture of Chi-Square Distributions

	Significance Level				
df (q)	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
0 vs. 1	1.64	2.71	3.84	5.41	6.63
1 vs. 2	3.81	5.14	6.48	8.27	9.63
2 vs. 3	5.53	7.05	8.54	10.50	11.97
3 vs. 4	7.09	8.76	10.38	12.48	14.04
4 vs. 5	8.57	10.37	12.10	14.32	15.97
5 vs. 6	10.00	11.91	13.74	16.07	17.79
6 vs. 7	11.38	13.40	15.32	17.76	19.54
7 vs. 8	12.74	14.85	16.86	19.38	21.23
8 vs. 9	14.07	16.27	18.35	20.97	22.88
9 vs. 10	15.38	17.67	19.82	22.52	24.49
10 vs. 11	16.67	19.04	21.27	24.05	26.07

This may work ok if only one new parameter is bounded … for example:

+ Random Intercept df=1: 2.71 vs. 3.84

+ Random Linear df=2: 5.14 vs. 5.99

+ Random Quad df=3: 7.05 vs. 7.82

Critical values such that the right-hand tail probability =

$0.5 \times \Pr(\chi^2_q > c) + 0.5 \times \Pr(\chi^2_{q+1} > c)$

Source: Appendix C (p. 484) from Fitzmaurice, Laird, & Ware (2004). *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley

Solutions for Boundary Problems when using −2ΔLL tests

- If adding random intercept variance only, use $p < .10$; $\chi^2(1) > 2.71$
  - Because $\chi^2(0) = 0$, can just cut *p*-value in half to get correct *p*-value

- If adding ONE random slope variance (and covariance with random intercept), can use mixture p-value from $\chi^2(1)$ and $\chi^2(2)$

$$\text{Mixture } p\text{-value} = 0.5 * \text{prob}(\chi_1^2 > -2\Delta\text{LL}) + 0.5 * \text{prob}(\chi_2^2 > -2\Delta\text{LL})$$

so critical χ^2 = 5.14, not 5.99

- However—using a 50/50 mixture assumes a diagonal information matrix for the random effects variances (i.e., it assumes the values for each are arrived at independently, which pry isn't the case)

- Two options for more complex cases:
  - Simulate data to determine actual mixture for calculating *p*-value
  - Accept that −2ΔLL is conservative in these cases, and use it anyway
    → I'm using ~ to acknowledge this: e.g., $-2\Delta\text{LL}(\sim 2) > 5.99$, $p < .05$

Missing Data in MLMs

- Common misconception is that MLM "handles" missing data, but there is a big difference in what happens to missing cases based on how variables are treated in model estimation
- Modern missing data techniques use the following vocabulary:
  - **Missing Completely at Random (MCAR):** probability of missingness is unrelated to value of missing responses (truly random mechanism)
  - **Missing at Random (MAR):** probability of missingness depends on the observed predictors or level 2 unit's other observed outcomes, but you can draw correct inferences after including (controlling for) these sources of observed data (conditionally random mechanism)
    - Full-information REML/ML and multiple imputation rely on this one—the shape of each group's likelihood function would be unaltered by adding back the missing obs
  - **Missing Not At Random (MNAR):** probability of missingness is indeed systematic but is not predictable based on the information you have (so everything will be some shade of wrong)

Univariate MLMs (MIXED in SAS, STATA, SPSS):

- Any case (row) with a **missing outcome is dropped** from the model
- In univariate MLMs, predictors are not part of the model likelihood—they are only relevant in computing each case's predicted outcome from the model fixed effects, so any case with a **missing predictor is dropped** from the model, too
  - Missing level-1 (person) predictors? Dropping these cases requires assuming MCAR for the person, but MAR-ish for the group (because the group still has other people still included)
  - Missing level-2 (group) predictors? Dropping these cases requires assuming MCAR for the group (because every case from that group will then be removed from the model)
- You may need to think about what predictors you want to examine PRIOR to model building, and pre-select your sample accordingly
  - Model fit statistics −2LL, AIC, and BIC are only directly comparable if they include the exact same observations (LL is sum of each height)
- Even if MCAR holds, dropping cases with missing data is a bad idea for its **reduction of statistical power**… so what to do instead?

Missing Data in MLMs

- Any cases that are missing model predictors (that are not part of the likelihood*) will not be used in that model
  - Bad for level-1 predictors (MCAR for persons; MAR-ish for groups)
  - Really bad for level-2 predictors (MCAR for whole group)

- Options for solving the problem of missingness:
  - *Bring the predictor into the likelihood (only possible in software for multivariate models, such as Mplus or SEM programs)
    - Its mean, variance, and covariances "get found" as model parameters
    - Predictor then has distributional assumptions (default multivariate normal), which may not be plausible for all predictors
  - Multiple imputation (and analysis of *each* imputed dataset)
    - Imputation also makes distributional assumptions!
    - Also requires all parameters of interest for the analysis model are in the imputation model, too (problematic for interactions or random effects)

Predicted Level-2 $\mathbf{U}_g$ Random Effects *(aka Empirical Bayes or BLUP Estimates)*

- Level-2 $\mathbf{U}_g$ random effects also require further explanation…
  - Empty two-level model: $\mathbf{y}_{pg} = \gamma_{00} + \mathbf{U}_{0g} + \mathbf{e}_{pg}$
  - U_{0g} values are deviated group means, right? Well, not exactly…

- 3 ways of representing size of individual differences in individual intercepts and slopes across level-2 groups:
  - Get each level-2 unit's OLS intercepts and slopes, save them to a dataset, and calculate their observed variance
  - Estimate variance of the $\mathbf{U}_g$'s (what we do in MLM)
  - Predict each group's $\mathbf{U}_g$'s; calculate their variance (2-stage MLM)

- Expected order of magnitude of variance estimates:
  - OLS variance > MLM variance > Predicted $\mathbf{U}_g$'s variance
  - Why are these different? "**Shrinkage**"

What about the U's?

- Level-2 unit $\mathbf{U}_g$ values are NOT estimated in the likelihood function
  - **G** matrix variances and covariances are sufficient statistics for the estimation process assuming multivariate normality of $\mathbf{U}_g$ values
  - Level-1 $\mathbf{U}_g$ random effects are **predicted** by asking for the SOLUTION on SAS RANDOM (pred without xb in STATA) as: $\mathbf{U}_g = \mathbf{G}_g \mathbf{Z}_g^T \mathbf{V}_g^{-1} (\mathbf{Y}_g - \mathbf{X}_g \gamma)$
    - Which then create individual estimates as $\boldsymbol{\beta}_{0g} = \gamma_{00} + \mathbf{U}_{0g}$ and $\boldsymbol{\beta}_{1g} = \gamma_{10} + \mathbf{U}_{1g}$

- What isn't obvious: the composite $\boldsymbol{\beta}_g$ values are weighted combos of the fixed effects ($\boldsymbol{\gamma}$) and level-2 OLS estimates (β_{OLS_g}) :

$$\text{Random Effects: } \boldsymbol{\beta}_g = \mathbf{W}_g \boldsymbol{\beta}_{OLS_g} + \left(\mathbf{I} - \mathbf{W}_g\right)\gamma \qquad \text{where: } \mathbf{W}_g = \mathbf{G}_g \left[\mathbf{G}_g + \sigma_e^2 \left(\mathbf{Z}_g^T \mathbf{Z}_g\right)^{-1}\right]^{-1}$$

  - The more true" variation in intercepts and slopes in the data (in **G**), the more the $\boldsymbol{\beta}_g$ estimates are based on level-2 unit OLS estimates
  - The more "unexplained" residual variation around the level-2 trajectories (in **R**), the more the fixed effects are heavily weighted instead
    - = **SHRINKAGE** (more so for groups with fewer persons, too)

What about the U's?

- Point of the story – **U_g** values are NOT single scores:
  - They are the mean of a distribution of possible values for each person (i.e., as given by the SE for each **U_g**, which is also provided)
  - These "best estimates" of the **U_g** values are shrunken anyway

- Good news: you don't need those **U_g** values in the first place!
  - Goal of MLM is to estimate and predict the variance of the **U_g** values (in **G**) with group-level characteristics directly in the model
  - If you want your **U_g** values to be predictors instead, then you need to estimate your model using multivariate MLM ("M-SEM")
  - We can use the predicted **U_g** values to examine potential violations of model assumptions, though… in SAS:
    - Get **U_g** values by adding: ODS OUTPUT SolutionR=dataset;
    - Get **e_{ti}** residuals by adding OUTP=dataset after / on MODEL statement
    - Add RESIDUAL option after / on MODEL statement to make plots

Estimation: The Grand Finale

- Estimation in MLM is all about finding the most likely estimates for the random effects variances and covariances
  - The more of them there are, the harder it is to find them (the more dimensions of the likelihood mountain there are to scale)
  - "Non-positive-definite" **G** matrix means "broken model"
  - Fixed effects are solved for after-the-fact in general MLMs, so they rarely cause estimation problems, but missing predictors can
  - Individual random effects are not model parameters, but can be predicted after-the-fact (but try never to use these as data)

- Estimation comes in two flavors:
  - ML → maximize the data; $-2\Delta LL$ to compare any nested models
  - REML → maximize the residuals; $-2\Delta LL$ to compare models that differ in their model for the variance ONLY